THE EMBO JOURNAL

Manuscript EMBO-2015-92759

Upstream ORFs act as pervasive translational repressors in vertebrates

Timothy G. Johnstone, Ariel A. Bazzini, Antonio J. Giraldez

Corresponding author: Antonio J. Giraldez, Yale University

Review timeline:

Submission date:	05 August 2015
Editorial Decision:	31 August 2015
Revision received:	28 November 2015
Editorial Decision:	04 January 2016
Revision received:	07 January 2016
Accepted:	08 January 2016

Editor: Anne Nielsen

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 31 August 2015

Thank you for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below.

As you will see from the reports, all three referees express interest in the findings reported in your manuscript; however they also raise a number of criticisms that you will have to address before they can support publication of the study in The EMBO Journal.

For such a revised manuscript I would particularly ask you to focus your efforts on the following points:

-> strengthen the functional insight on the contribution from uORF regulation as requested by ref#1
-> extend the description of the analysis to more clearly explain the basis for annotations and conclusions (ref #2). In addition, this referee lists a number of suggestions for additional analysis that could be included to strengthen the argument for uORF regulatory potential.
-> in addition, I would ask you to comment/clarify the many minor issues related to data presentation and analysis description pointed out by all three referees.

Given the referees' overall positive recommendations, I would thus invite you to submit a revised version of the manuscript, addressing the comments of all three reviewers. I should add that it is EMBO Journal policy to allow only a single round of revision, and acceptance of your manuscript will therefore depend on the completeness of your responses in this revised version.
Thank you for the opportunity to consider your work for publication. I look forward to your revision.

REFEREE REPORTS

Referee #1:

The authors describe a computational analysis of previously published Ribo-seq data in zebrafish, human and mouse, focused on the analysis of translated upsteam and overlapping ORFs (uORFs and oORFs) in mRNA genes. The manuscript is very well written and most of the statistical analysis is sound. The analyses pertaining to conservation of uORFs and avoidance of uORFs and oORFs in UTR are particularly interesting and compelling. While most of the discussed concepts are not new, overall this the most comprehensive computational analysis of this topic to date. However, without experimental validation of causality, most of the analyses point at correlations between sequence features and observed translation events in 5'UTRs and translation efficiencies of the main ORFs, rather than proving repressive effects of the translation events, limiting the overall significance of the results.

Major comments:
1. Throughout the manuscript, the authors mainly report correlations, but use causal terms, without performing any experiments that show that the translation events in 5'UTRs are indeed directly causing repression. Therefore, the authors should either perform validation experiments (by using constructs with different uORF configurations they work with in 5'UTRs and show how those affect expression of GFP/luciferase in the zebrafish embryo) or rewrite the manuscript to avoid any misleading statements. For example, in the abstract there are statement like "uORFs act primarily as potent regulators of translation and RNA levels, with a similar repressive capacity to miRNAs". This sentence suggests that there is evidence that uORFs globally regulate translation, whereas the authors only report a correlation between uORF presence and lower translation efficacy, without any direct evidence of a "regulation" taking place. The data from mESCs and mEBs in Fig. 3E-F is far from sufficient for addressing this, as the two cell lines are so different and any differences in translation are probably indirect. In the next sentence in the abstract the authors claim "We identify uORF number, ... as the features with the strongest influence on translation." - again, there is no evidence that these features influence translation, only that they are correlated with lower translation efficiency. Throughout the paper there are many statements like this that are implying causality (basically any place where the authors say the uORFs "repress" translation), and they are too numerous to list here. Also particularly notable is the statement in the Discussion "the predominant function of uORFs appears to lie in their broad regulatory role as translational repressors.". How can this be claimed based on the results in the manuscript? No direct function was tested, and certainly no alternatives to functions as translation repressors were explored. Unless the authors provide experimental evidence for causality all of these statements throughout the manuscript need to removed or rephrased.

2. Some of the observed correlation between presence of uORFs in 5'UTRs and lower translation efficiencies and lower mRNA levels could result from 5'UTR length and composition effects, that are not controlled for (particularly because highly expressed and efficiently translated genes typically have short 5'UTRs). The authors need to show that uORF-containing mRNAs are repressed compared to 5'UTR length- and composition-matched control mRNAs that do not have translated uORFs (in particularly for Figure 2D-E, Fig. 3A-F).

Minor comments:
1. PhyloCSF analysis should also be performed on the mammalian uORFs to test if the results are consistent between zebrafish and mammals.

2. If the uORF translation indeed affects mRNA level through NMD, it can be tested by inhibition of NMD (e.g., by targeting Upf1) followed by testing of mRNAs levels of uORF-containing mRNAs (compared to those without uORFs).

3. Are the transcripts that exhibit enrichment in uORFs in human, mouse, zebrafish enriched for orthologs of each other? I.e., is the feature of having multiple translated uORF conserved across vertebrates?

4. The analysis in the "uORF regulation is conserved in vertebrates" section should be done also for

human-zebrafish and mouse-zebrafish comparisons. The results will be of interest whether the authors observe conservation or not.

5. Fig 4D does not appear to be referred to in the text, and its not clear what it shows.

Referee #2:

This paper uses ribosome profiling data generated in a previous study published by this same group in EMBO J in which uORF translation was detected in the zebrafish translatome analyzed at different developmental stages. That paper presented a strategy for identifying uORFs that exploits the codon phasing of high-resolution ribosome footprints derived from 28-29 nt-ribosome protected fragments (RPFs), leading to identification of several hundred high-confidence uORFs initiated by AUG start codons. In this paper they analyze other properties of the uORFs and the genes containing them in zebrafish, and also of uORFs identified in published datasets for mouse and humans, in an effort to uncover the predominant functions of uORFs in regulating translation and mRNA turnover. Using their algorithm called ORFscore, they identify 3875 high- or medium-confidence translated uORFs in zebrafish, which show good phasing of ribosomes in the correct reading frame, and another 3043 uORFs that lacked good phasing but contained characteristic peaks/double peaks at the AUG and stop codons, respectively, which represents 45% of all annotated transcripts. Another 969 transcripts contain an uORF that overlaps the main coding sequence (dubbed an oORF). Using PhyloCSF, they find that only a very small proportion of the uORFs (<1%) displayed evidence for conservation of the peptide sequence in different fishes, suggesting that the vast majority do not function in translational control mechanisms in which the uORF peptide arrests elongation and blocks scanning to the downstream start codon. They obtain evidence however that uORFs and oORFs both tend to repress translation of downstream CDS by other means, by finding that the presence of uORFs is associated with lower average translational efficiencies (TE), with greater numbers of uORFs in the same mRNA being associated with relatively lower average TEs, at all developmental stages. The same trend was observed for uORFs identified in published datasets for the mouse and human translatomes, for various different types of mouse or human cells. They argue that uORF translation is causative and not merely associated with reduced TEs based partly on the observation in Fig. 3E that the association of uORFs with reduced TE is weaker in murine embryoid bodies versus other mouse/human cells analyzed and citing previous observations that uORF translation is relatively reduced in embryoid bodies. They also present evidence that uORF features that would favor their translation or diminish the probability of reinitiation at the downstream CDS appear to be selected against in evolution, which should mitigate the potential inhibitory effects of the uORF and render it subject to modulation to effect quantitative translational control. They also present evidence that the relative TEs of uORFs and CDS are often conserved across species, consistent with the conservation of translational control by uORFs. They conclude that uORFs generally function as conserved translational repression elements.

General Critique:

This paper points out some interesting features of uORFs and relationships between the presence of uORFs and the TEs of the main CDS in mRNAs consistent with a general and widespread function of uORFs in dampening translation and reducing mRNA stability; and suggestive of selection against various features that would enhance translation of the uORFs or reduce the frequency of reinitiation following uORF translation in order to mitigate their inhibitory effects, including multiple uORFs, optimal uORF start codon context, long uORF length, and proximity to the main CDS. However, there are many uncertainties about exactly how the analyses were conducted, and there are a number of obvious analyses that could have been included that would potentially provide stronger evidence for the main conclusions. In Bazzini et (2014), this group established a rigorous method of uORF discovery, which they seem to have largely abandoned, with no explanation, in selecting sets of uORF for further analyses, and it's unclear whether the uORFs analyzed in mouse/human cells using datasets published by other groups were subjected to even the most minimal validation of translation (peaks at start and stop codons). It seems reasonable to request that the various analyses of uORF properties and their association with TEs of the main CDS be conducted using the uORF sets designated as high-/medium-confidence translation in addition to all predicted uORFs or all uORF sets that include low-confidence uORFs. Second, finding associations between uORF numbers and reduced average TEs for sets of genes is interesting, but it would be

much more compelling to find an inverse relationships between the TEs of uORFs and the TEs of downstream CDS, as the impact of an uORF on downstream translation will be governed by the fraction of scanning ribosomes that are waylaid by the uORF start codon versus continued scanning to the main CDS. When there are multiple uORFs, the uORF with the highest TE will be the most efficacious as a negative control element. Even better than establishing an inverse relationship between uORF and CDS TEs in each different cell type would be to show that changes in TE of uORFs are associated with inverse changes in TE of the downstream CDS when comparing the same cell types in different developmental states. This could be accomplished by considering changes in TE that occur in different stages of zebrafish development, or between murine embryoid bodies and MEFs. Along the same lines, one would like to know whether the TE of uORFs is inversely correlated with uORF length and directly correlated with distance from the downstream CDS, as would be expected if well-translated uORFs are optimized for reinitiation (by short lengths and large intercistronic distances). A related question is whether the uORF TEs are highly correlated with the uORF initiation context score, since the latter is being used as a proxy for uORF TEs in Fig. 4B. A particularly puzzling feature of the findings in Figs.4A-5 is that the TE ratios of uORFs to CDS are generally greater than one (Fig. 5) but the uORF AUG context scores are very low compared to the CDS scores. This seems possible only if most genes have multiple uORFs whose ribosome occupancies are being combined to calculate the 5'UTR TE, but this can be discounted using the pie-charts in Fig. 1. In summary, a much better description of the analyses they conducted is required, the analyses should be repeated where necessary on medium-/high-confidence uORFs, and additional analyses are required that correlate the TE of uORFs with CDS TEs and uORF properties. Even better would be to analyze changes in uORF TEs versus changes in CDS TEs between different cell types or developmental stages.

Specific comments (Some of the comments below repeat or elaborate points made above in the General critique):

-the number of uORFs described in zebrafish in this paper is far greater than that described previously (Bazzini et al 2014), even for just the high- and medium-confidence uORFs: 3875 vs several hundred in the 2014 study. This seems to occur because they did not apply the "coverage" criterion here that was described in the previous paper, which demands that a large fraction of the codons in the uORF exhibit at least one read in the correct frame for each codon. It seems important to justify the decision to abandon in the present study the "coverage" criterion, which seemed very reasonable; and also to include low-confidence translated uORFs (just a peak at AUG/stop) in their analyses of uORF properties and impact on translation.

-the low confidence uORFs pass neither the ORFscore or coverage criteria and seem to exhibit only peaks at the start and in-frame stop codons. They should describe how high these peaks needed to be relative to the surrounding ribosome density to qualify as indicators of a translated uORF.
-Related to the previous comment, it seems likely that in many of their analyses of the properties of uORFs and uORF-containing mRNAs, they have included the low-confidence translated uORFs, and in some cases all predicted uORFs whether or not there is strong evidence that they are translated. It needs to be stated clearly what uORF set is being interrogated in each analysis. And it seems important to repeat each analysis for only the high and medium confidence translated uORFs in situations where this was not done, to determine whether the same conclusions hold.

-Fig. 2A-B: It's unclear how the conservation of uORF sequences in different fish species was determined, notwithstanding the explanation given in the Methods. Can it be stated that the 27 sequence-conserved uORFs are contained in the 5'UTRs of the paralogous genes in the different fish species? Showing some typical alignments of conserved uORFs in the different species, annotated with transcription start sites and the beginning of the CDS, in a Supplementary figure should be required. Also, it's unclear why they interrogated only 8225 of the possible 22,138 predicted uORFs in Fig. 2A, and only 755 of translated uORFs out of 3875 high-/medium-confidence uORFs in panel B. Much more explanation is required here about which uORFs were interrogated and how it was done.

-Fig. 2D-E and G-H & and Fig. S2A-H: These are situations where it's unclear whether all potential uORFs, only low/medium/high confidence translated uORFs, or only medium/high-confidence translated uORFs have been analyzed here. Based on their statements, it seems that all predicted uORF-containing transcripts were analyzed for the association of uORFs with lower average TEs.

Thus, it's important to determine if the effect is even stronger when considering only medium/high-confidence translated uORFs.

-p. 6, bottom para.: it's unclear whether the uORFs identified in the human and mouse cell datasets were subjected to the same analysis of reading frame phasing using only 28-29 nt reads to identify high/medium confidence uORFs, whether a requirement only for a peak at the AUG and stop codon was imposed, or whether all predicted uORFs were analyzed without any attempt to obtain evidence for their being translated. Much more detail is required in the description of how they did their analyses of these published datasets.

-last sentence of p. 6, and lines 6-7 on p.7: the authors should be more careful in interpreting the association between the presence of uORFs and reduced average TEs for different gene sets. This association is consistent with uORFs exerting translational repression but certainly does not demonstrate it in the manner implied in these two sentences. The last sentence in the section on uORF abundance correlating with mRNA levels on p. 7 is similarly too strong.

-Fig. 3E: It seems important to use a statistical test to establish that the association of uORFs with lower TE values in embryoid bodies is significantly weaker than the correlations observed in the other mouse or human cell lines shown in Fig. 3 and Fig. S2. Also, in lines 14-16 on p. 7, the results in Fig. 3E-F should be compared to all others in Fig. 3 and Fig. S2 to make this point.

-Fig. 3E: It also seems important to verify that the TE of the set of uORFs they identified using their uORF-discovery approaches in embryoid bodies is actually lower than the TE of uORFs identified in other mouse or human cells, rather than just citing this fact from Ingolia et al (2011) as it's unclear how much overlap there might be between the uORFs interrogated in Ingolia et al (2011) and those examined here.

-The embryoid data is being used to establish that the level of uORF translation is inversely related to the TE of the CDS. A more compelling argument could be made by identifying zebrafish uORFs that exhibit changes in TE during development and determine whether the downstream CDS exhibit inverse changes in TE in the manner expected if uORFs are generally repressive. If uORFs generally don't change their TEs during development, then they should ask whether there is an inverse correlation between the TE of the uORF and the TE of the downstream CDS in different zebrafish, mouse and human cells, as it is not the number of uORFs but rather the TEs of the uORFs that should most strongly determine how much they repress downstream initiation by diverting scanning ribosomes from reaching the AUG for the CDS. They could also ask whether there is an inverse correlation between the changes in TE in uORFs and changes in TE in the downstream CDS when comparing mouse embryoids to MEFs. Similarly, the conclusion that uORFs generally decrease mRNA stability would be more convincing if changes in uORF TEs during development were inversely correlated with changes in mRNA levels, or if the TE of uORFs (not just uORF number) was inversely correlated with mRNA levels. Also, if uORF translation evokes reduced mRNA levels, then the negative correlation between uORFs and RNA levels shown in Fig. 2G should be significantly less apparent in mouse embryoid bodies vs mESCs or MEFs.

-last 2 lines of p. 7: this is confusing. Fig. S4 shows no correlation between TLS length and the CDS TE for mRNAs lacking uORFs. Presumably, this allows them to rule out the possibility that mRNAs with more uORFs have a lower TE because they tend to have longer 5'UTRs/TLSs. Is the latter true, ie. do mRNAs with more uORFs tend to have longer TLSs?.

-p. 8, first full para.: It's hard to be convinced that oORFs are associated with lower TEs than uORFs because they have grouped together transcripts with 1 or 2 oORFs and are comparing the results to transcripts with 1, 2, or 3 uORFs. The proper comparison would be to transcripts with 1 or 2 uORFs, with a statistical test to show a stronger anticorrelation for the oORFs vs. uORFs.

-p. 8 and Fig. S5B: The correlation shown in Fig. S5B is very small in magnitude even if highly significant, questioning the biological importance of intercistronic distance. Given the strong theoretical basis for the effect of intercistronic distance on reinitiation from the work of Kozak and Hinnebusch/Wek (on GCN4/ATF4), this analysis should be repeated with the high/medium-confidence translated uORFs; and they should also ask whether intercistronic distance is directly correlated with the TE of the uORF as would be expected if large distances are being selected for

during evolution to mitigate the inhibitory effects of well translated uORFs by allowing high levels of reinitiation.

-p. 8 text and Fig. S5C: Given the lack of correlation between uORF length and TE of the CDS, it also seems important to re-do this analysis considering only the high-/medium-confidence uORFs, and also to ask whether uORF length is inversely correlated with the TE of the uORF as would be expected if long uORF length is being selected against in well translated uORFs to permit higher levels of reinitiation.

-Fig. S5: Although there is a significant negative correlation here, it is quite weak. It should be verified that this is true even if high-/medium-confidence translated uORFs are analyzed.
-p. 9, top para. and Fig. 4A: Given that the ORF AUG score is so much lower for oORFs versus uORFs, yet oORFs appear to be more inhibitory, they should determine whether the TE of oORFs is also generally lower than that of uORFs. If so, the relatively greater inhibition by oORFs, despite being translated by fewer scanning ribosomes, would derive from the fact that they overlap the main CDS in a way that precludes reinitiation and also might interfere with initiation complex assembly at the CDS AUG codon.

-Fig. 4C: the schematics are hard to fathom and should be explained in the legend.

-Fig. 4F and text on p. 9: Again it's unclear whether all potential uORFs are examined for mean length, and whether the conclusion also holds for high/medium confidence translated uORFs, and also whether the TE of the uORF is inversely correlated with its length as might be expected if short uORF lengths are being selected during evolution for well-translated uORFs? It's not explained where the cited figure of <60nt (presumably 41.5 nt) for the mean uORF length in all vertebrate species comes from-which species, which types of uORFs considered? They also don't indicate the mean uORF length obtained in the TLS shuffle simulation-how much larger than 41.5 nt is it?. Finally they don't discuss any mechanistic basis for the apparent selection pressure against uORF length.

-bottom of p. 9/top of p. 10: cite Fig. 4D versus 4C.

-p. 10, line 5: context data for human/mouse uORFs should be shown in Supplemental Figs.

-Fig. 5B-D: first, this figure was not cited in the text. Second, I assume that log(2) ratios are being plotted. Third, it is surprising that the ~average UTR/CDS TE ratio generally appears to be ~2, indicating that there are typically 4-fold greater ribosome densities in 5'UTRs vs.CDS. This would be surprising, considering the typical genes shown in Fig. 1D where uORF densities are less then CDS densities, and the fact that most genes don't contain multiple uORFs and uORF AUG context is so poor.

-Fig. 5B-D and text on p. 10, 1st full para. It's not clear that the differences in correlation coefficients between panels B, C,D are significantly different from one another. Perhaps a better analysis would be to make notched box-plots of the ratios in the different cell types and determine whether the notches fail to overlap for the comparisons in C and D but not for panel B.

-p. 10, line 6 from bottom: stipulate what kind of uORFs is referred to here-I believe it's high/medium confidence translated uORFs, but not sure for the oORFs. Also what kind of uORFs is referred to a few lines down in citing 61% of zebrafish uORFs being translated?

-p. 12, lnes 12-14: This is an overstatement. They have shown a reasonably strong correlation between the relative TEs of uORFs versus downstream CDSs across species. This is consistent with, but certainly does not prove that regulation by particular uORFs is conserved across species. A few lines down, it is also incorrect that uORFs have sequence contexts less favorable than by chance, which is true only of oORFs.

Referee #3:

The location and impact of uORFs is an important question relevant to many biological processes including stress response (in human ATF4) and nutrient regulation (yeast GCN2 and plant bZIP11). Here, Johnstone and colleagues search 5' untranslated regions for potential upstream open reading frames (uORFs), and what they term overlapping upstream open reading frames (oORFs), using in silico prediction. These candidate uORFs and oORFs are then assessed as sites of actual translation initiation using the authors' software, ORFScore, which utilizes the nucleotide level framing information provided by ribosome profiling. They performed this analysis in zebrafish, mouse, and human data sets, showing the pervasive appearance of u/oORFs in transcription leader sequences (TLSs). The authors then study the roles of u/oORFs through correlation analysis of u/oORF appearance, number, and feature with either translational efficiency (TE: ribosome protected fragments/mRNA seq) or overall mRNA abundance. This TE analysis repeatedly indicates u/oORFs are acting as translational repressors-moreover, the number, proximity, and context of the u/oORF plays a role in determining the severity of translation repression. Analysis of mRNA abundance suggests additional roles or mechanisms for u/oORFs in post-transcriptional regulation, as u/oORF-containing transcripts tend to have lower abundance than transcripts without. While the actual u/oORF sequences may not be evolutionarily conserved, the authors argue that the presence and impact of these features are conserved across zebrafish, mouse, and human.

The annotation of uORF boundaries here builds on previous work (as noted in several of the cited sources) to contribute to the field by determining the location of translated uORFs using ribosome profiling data. The effect of u/oORFs on the translation and stability of mRNAs remains a topic of debate, and the improved u/oORF annotation presented here allows the authors to correlate truly translated u/oORFs with other mRNA properties. I thus think the paper will attract substantial interest, and I support the publication of this paper provided the following concerns can be addressed.

Specific major concerns

1. Metagene analysis can be heavily biased by genes with a particularly strong signature- a distribution of values compiled for the metagene scores is a useful supplement to confirm the effect isn't driven by a single, large effect gene.

2. The authors describe TE/repression as difficult to study in cases of oORFs because of overlapping regions/counts. This is can be circumvented using computationally defined boundaries that exclude the region of overlap: for the CDS downstream the of the stop codon in the oORF, and for the oORF, upstream of the CDS. Moreover, in cases where the CDS and oORF do not share a frame, information can be gained even from this region of overlap by looking at the framing data itself rather than regional counts.

3. The use of CAU context versus AUG context for TLS background estimation

Minor concerns

5. Per the reliance of this paper on several previously published papers (e.g. ORFScore, PhyloCSF), a brief description of them would be of considerable value for assaying the merit of the work.

6. In 4B, there is potential for ascertainment bias - a good AUG context may confer better translation, and this may in turn better regulate downstream CDS TE, but also ensures higher confidence detection of the uORF. It would be useful to compare AUG context in high versus low confidence uORFs, and then compare how these affect downstream TE.

7. 4D isn't discussed in the text. There is a discussion of potentially similar data, but refers to an unrelated cartoon figure 4C.

8. Several issues in figure or legend clarity:
a. 1A lacks adequate labeling per color coding (orange v grey v purple) and has some difficult to distinguish colors (low v med v high) and ribosomes (med v high). The legend is also

unclear/misleading in describing a subset of uORFs as containing oORFs.

b. the arrangement of 1C is non intuitive given the way these are usually presented (left to right as the standard vs the up to down for 5' to 3'). Moreover, the framing data in Low Confidence uORFs is described as potentially due to overlapping frames, though there's no clear evidence in this figure to support that.

c. 1D is nigh illegible because of size; moreover, the CDS isn't depicted in a graphically meaningful way. This makes it difficult to judge the size, location, and relevance of the uORFs. Additionally, "mRNA Seq" or RPKM is a more informative title axis than simply "Input".

d. fig 3 changes in axes change potential magnitude of effect

e. 3C,D have no definition of LIF in the figure, supplement, or text-if the absence or presence is worth noting, it is an acronym worth defining. This figure also uses a different x axis than 3A,B,E,F. Changes in the x-axis on cumulative distributions can be misleading for magnitude of effect. If all six graphs are to be compared to each other to remark on differences or similarities between species, the axes need to be the same.

f. 4A's x axes is not intuitively clear (how you would grade a better context) without having read Grzegorski.

g. 4C is a graphical abstract missing enough pieces of data to actually clarify what it is intending to show.

h. Figure 5E has r=.237 and is considered no correlation, whereas S5B has r=.147 and is considered correlated.

i. S5C,D lacks legend descriptions.

1st Revision - authors' response 28 November 2015

First, we would like to thank you for your valuable suggestions and remarks, your input demonstrates keen interest and knowledge of the field and has been very constructive in refining the paper.

We have been able to address most of the comments and analyses you suggested. In particular, as highlighted in our correspondence with the editor at the time of receiving reviews, we have:

-> Strengthened the functional insight on the contribution from uORF regulation as requested: by generating reporters to testing the hypothesis put forward in the paper demonstrating causality of the regulation *in vivo*, and by performing a metanalysis across 23 different samples of uORF-mediated translation repression.

-> Extended descriptions of the analysis to more clearly explain the basis for annotations and conclusions (ref #2). We have also made an effort to address the comments made by this reviewer with the goal of strengthening the argument for uORF regulatory potential.

-> Addressed/commented the minor points raised by all three of the referees.

Below you can find our answers to the comments on the submitted manuscript entitled: “Upstream ORFs act as prevalent translational repressors in vertebrates.”
The manuscript has been revised taking into account the remarks raised, which we have enumerated below.

We hope that our answers address the concerns raised, and that the manuscript is now acceptable for publication. We look forward to receiving your reply.

Referee #1:

Major comments:
1. Throughout the manuscript, the authors mainly report correlations, but use causal terms, without performing any experiments that show that the translation events in 5'UTRs are indeed directly causing repression. Therefore, the authors should either perform validation experiments (by using constructs with different uORF configurations they work with in 5'UTRs and show how those affect expression of GFP/luciferase in the zebrafish embryo) or rewrite the manuscript to avoid any misleading statements. For example, in the abstract there are statement like "uORFs act primarily as potent regulators of translation and RNA levels, with a similar repressive capacity to miRNAs". This sentence suggests that there is evidence that uORFs globally regulate translation, whereas the authors only report a correlation between uORF presence and lower translation efficacy, without any direct evidence of a "regulation" taking place. The data from mESCs and mEBs in Fig. 3E-F is far from sufficient for addressing this, as the two cell lines are so different and any differences in translation are probably indirect. In the next sentence in the abstract the authors claim "We identify uORF number, ... as the features with the strongest influence on translation." - again, there is no evidence that these features influence translation, only that they are correlated with lower translation efficiency. Throughout the paper there are many statements like this that are implying causality (basically any place where the authors say the uORFs "repress" translation), and they are too numerous to list here. Also particularly notable is the statement in the Discussion "the predominant function of uORFs appears to lie in their broad regulatory role as translational repressors.". How can this be claimed based on the results in the manuscript? No direct function was tested, and certainly no alternatives to functions as translation repressors were explored. Unless the authors provide experimental evidence for causality all of these statements throughout the manuscript need to removed or rephrased.

Thank you for pointing this out. We have also adjusted the language of the manuscript to avoid unsupported statements of causality. Additionally, we have improved the analysis linking translation to repression – we have added many additional cell/tissue types and show that across samples, uORF translation is linked to repression (Fig. 3E). Most importantly, we have undertaken reporter experiments in zebrafish using 5 different configurations of uORFs, oORFs and AUG context (included in Figure 5) addressing causality and relative strength of the regulation by uORFs/oORFs.

2. Some of the observed correlation between presence of uORFs in 5'UTRs and lower translation efficiencies and lower mRNA levels could result from 5'UTR length and composition effects, that are not controlled for (particularly because highly expressed and efficiently translated genes typically have short 5'UTRs). The authors need to show that uORF-containing mRNAs are repressed compared to 5'UTR length- and composition-matched control mRNAs that do not have translated uORFs (in particularly for Figure 2D-E, Fig. 3A-F).

We have analyzed the relationship between 5'UTR length and translation and find that this cannot explain that more uORFs cause stronger translational repression. We have shown in supplementary figure S4A,B that when controlling for the number of uORFs, longer TLSs in fact have a positive or neutral effect on translation efficiency. Since we do not observe any negative correlation between 5'TLS length and translation efficiency, we reason that UTR length is not responsible for the observed reduction in translation in the uORF-containing group. As for composition, we have checked in zebrafish and do not observe a meaningful difference between the GC content of 5'TLSs containing a uORF vs. 5'TLSs lacking uORFs (0.451, 0.456 respectively).

Minor comments:

1. PhyloCSF analysis should also be performed on the mammalian uORFs to test if the results are consistent between zebrafish and mammals.

We have performed the phyloCSF analysis on mammalian ORFs and observe that, consistent with zebrafish, only a very small proportion of uORFs show evidence for conservation. These results have been added in the text and supplemental figure S7.

2. If the uORF translation indeed affects mRNA level through NMD, it can be tested by inhibition of NMD (e.g., by targeting Upf1) followed by testing of mRNAs levels of uORF-containing mRNAs (compared to those without uORFs).

We appreciate this suggestion, and it is a great set of experiments that would no doubt yield interesting results, but we see this as a follow-up project beyond the scope of the current paper. We have modified the text to leave open other mechanisms of regulation.

3. Are the transcripts that exhibit enrichment in uORFs in human, mouse, zebrafish enriched for orthologs of each other? I.e., is the feature of having multiple translated uORF conserved across vertebrates?

We have now performed this analysis and observed that there is a significant correlation of the ORF number z-score between one-to-one homologs. This has been added to the text and methods section. 31% of highly uORF-enriched mouse transcripts (Z-score > 2) with a one-to-one orthology mapping between human and mouse were also significantly enriched in human.

4. The analysis in the "uORF regulation is conserved in vertebrates" section should be done also for human-zebrafish and mouse-zebrafish comparisons. The results will be of interest whether the authors observe conservation or not.

We have tried performing this analysis, but it was not tractable to compare zebrafish data with data from human/mouse due to the following caveats: (1) because of the teleost genome duplication, it is difficult to map orthology between zebrafish and mammals, so due to high paralogy in the zebrafish genome the set of orthologous genes is too small to support a strong conclusion and (2) zebrafish ribosome profiling data are whole-embryo, whereas data in other organisms is tissue-specific, and we do not feel it fair to directly compare aggregate translation levels with those from a specific tissue.

5. Fig 4D does not appear to be referred to in the text, and its not clear what it shows.
The figure is now properly referenced in the text.

Referee #2:

General Critique:

This paper points out some interesting features of uORFs and relationships between the presence of uORFs and the TEs of the main CDS in mRNAs consistent with a general and widespread function of uORFs in dampening translation and reducing mRNA stability; and suggestive of selection against various features that would enhance translation of the uORFs or reduce the frequency of reinitiation following uORF translation in order to mitigate their inhibitory effects, including multiple uORFs, optimal uORF start codon context, long uORF length, and proximity to the main CDS**.** However, there are many uncertainties about exactly how the analyses were conducted, and there are a number of obvious analyses that could have been included that would potentially provide stronger evidence for the main conclusions. In Bazzini et (2014), this group established a rigorous method of uORF discovery, which they seem to have largely abandoned, with no explanation, in selecting sets of uORF for further analyses, and it's unclear whether the uORFs analyzed in mouse/human cells using datasets published by other groups were subjected to even the most minimal validation of translation (peaks at start and stop codons). It seems reasonable to request that the various analyses of uORF properties and their association with TEs of the main CDS be conducted using the uORF sets designated as high-/medium-confidence translation in addition to all predicted uORFs or all uORF sets that include low-confidence uORFs.

This is addressed in later comments

Second, finding associations between uORF numbers and reduced average TEs for sets of genes is interesting, but it would be much more compelling to find an inverse relationships between the TEs of uORFs and the TEs of downstream CDS, as the impact of an uORF on downstream translation

will be governed by the fraction of scanning ribosomes that are waylaid by the uORF start codon versus continued scanning to the main CDS. When there are multiple uORFs, the uORF with the highest TE will be the most efficacious as a negative control element. Even better than establishing an inverse relationship between uORF and CDS TEs in each different cell type would be to show that changes in TE of uORFs are associated with inverse changes in TE of the downstream CDS when comparing the same cell types in different developmental states. This could be accomplished by considering changes in TE that occur in different stages of zebrafish development, or between murine embryoid bodies and MEFs.

This is addressed in later comments

Along the same lines, one would like to know whether the TE of uORFs is inversely correlated with uORF length and directly correlated with distance from the downstream CDS, as would be expected if well-translated uORFs are optimized for reinitiation (by short lengths and large intercistronic distances).

Thank you for suggesting this additional analysis. Indeed, uORF TE is inversely correlated with uORF length (r=-0.22, p=1.0e-92, 5hpf), however we do not observe significant correlation between intercistronic distance and uORF TE (r=0.002, p=0.83, 5hpf). These results have been included in the manuscript.

A related question is whether the uORF TEs are highly correlated with the uORF initiation context score, since the latter is being used as a proxy for uORF TEs in Fig. 4B.

uORF translation efficiencies are significantly correlated with the uORF initiation context score (r=0.20, p<2.2e-16), moreso in single uORF transcripts (r=0.30, p<2.2e-16) – we suspect this difference between single uORF transcripts and multi-uORF transcripts is due to confounding translation effects of early uORFs on the translation of subsequent/overlapping uORFs. This plot has been included in supplementary figure S6.

A particularly puzzling feature of the findings in Figs.4A-5 is that the TE ratios of uORFs to CDS are generally greater than one (Fig. 5) but the uORF AUG context scores are very low compared to the CDS scores.

Though uORF AUG contexts are unfavorable compared to the CDS, they are still the first AUG encountered by the ribosome, and thus will still be engaged with some baseline frequency. Additionally, many 5' UTRs are short, and the RPF contribution from multiple uORF initiation and termination peaks contributes much more translation signal when calculating the TE of a whole, short 5' UTR as opposed to the translation within the internal codons of a CDS.

Specific comments (Some of the comments below repeat or elaborate points made above in the General critique):

-the number of uORFs described in zebrafish in this paper is far greater than that described previously (Bazzini et al 2014), even for just the high- and medium-confidence uORFs: 3875 vs several hundred in the 2014 study. This seems to occur because they did not apply the "coverage" criterion here that was described in the previous paper, which demands that a large fraction of the codons in the uORF exhibit at least one read in the correct frame for each codon. It seems important to justify the decision to abandon in the present study the "coverage" criterion, which seemed very reasonable; and also to include low-confidence translated uORFs (just a peak at AUG/stop) in their analyses of uORF properties and impact on translation.

Due to their small size (even compared to the previously analyzed set of lincRNA ORFs), the coverage criterion has little effect on translation calls for the majority of uORFs - a uORF with 10 internal codons would require only a single codon to be covered in order to pass the coverage filter – thus it was not taken into account. Use of more updated transcript annotation sets has also improved the quality of 5'TLSs and allowed us to find additional uORFs that pass our ORFScore threshold. We feel that the medium and low confidence sets are justified, given the clear signatures of translation we observe at a metagene level (Fig. 1C) – the confidence levels reflect detectability by our classification criteria, but not necessarily a gradient of translation levels, since the ribosome

profiling is strongly influenced by the level of expression of that mRNA. Additionally, we feel that inclusion of uORFs not designated as 'high confidence' is justified due to the fact that we observe effects on translation and RNA levels indistinguishable from those of the high confidence set, as is now included in Supplementary Figure S1 (I&J).

-the low confidence uORFs pass neither the ORFscore or coverage criteria and seem to exhibit only peaks at the start and in-frame stop codons. They should describe how high these peaks needed to be relative to the surrounding ribosome density to qualify as indicators of a translated uORF.

We did not use these peaks to classify ORF translation, but rather to validate post-hoc that our classification criteria for the low-confidence translated uORFs were indeed selecting translated regions. We believe based on our observations that even the low confidence detected ORFs are translated, but in many cases it is likely that the lack of frame is due to low RNA expression or because there are several uORFs that overlap with one another being translated for those transcripts, therefore reducing the frame bias. Indeed this is apparent because, at a metagene level, once the stop codon of low confidence uORFs is reached there is a still a phasing signal out-of-frame after the stop codon(Figure 1C, low confidence, green bars after stop codon)

-Related to the previous comment, it seems likely that in many of their analyses of the properties of uORFs and uORF-containing mRNAs, they have included the low-confidence translated uORFs, and in some cases all predicted uORFs whether or not there is strong evidence that they are translated. It needs to be stated clearly what uORF set is being interrogated in each analysis. And it seems important to repeat each analysis for only the high and medium confidence translated uORFs in situations where this was not done, to determine whether the same conclusions hold.

We have included additional language more clearly stating which sets of uORFs were used in each analysis. We have also performed various analyses, notably the repression analyses, with only high/medium confidence uORFs, as is addressed in other comments.

-Fig. 2A-B: It's unclear how the conservation of uORF sequences in different fish species was determined, notwithstanding the explanation given in the Methods. Can it be stated that the 27 sequence-conserved uORFs are contained in the 5'UTRs of the paralogous genes in the different fish species?

We have added more detailed language about PhyloCSF and have now included a new Supplementary Figure S8, which shows representative alignments for conserved, weakly conserved, and non-conserved uORFs. Though it is probable that the 27 sequence-conserved uORFs are contained in paralogous 5'UTRs given the nature of how the multiple alignments are made, we cannot ensure that all sequences present in the multiple alignment are in functionally orthologous regions in all species due to the lower quality of transcriptome assemblies in the other fish species.

Showing some typical alignments of conserved uORFs in the different species, annotated with transcription start sites and the beginning of the CDS, in a Supplementary figure should be required.

We have added representative alignments in Figure S8.

Also, it's unclear why they interrogated only 8225 of the possible 22,138 predicted uORFs in Fig. 2A, and only 755 of translated uORFs out of 3875 high-/medium-confidence uORFs in panel B. Much more explanation is required here about which uORFs were interrogated and how it was done.

Language has been added to the methods section to clarify the phyloCSF analysis. The primary contributor to the reduced number of analyzed uORFs relative to the whole pool is uORF length, as the minimum length for our phyloCSF analysis is 10AA (which is smaller than the default length of 15AA), the limit at which it is possible to make reasonably confident conservation calls.

-Fig. 2D-E and G-H & and Fig. S2A-H: These are situations where it's unclear whether all potential uORFs, only low/medium/high confidence translated uORFs, or only medium/high-confidence translated uORFs have been analyzed here. Based on their statements, it seems that all predicted uORF-containing transcripts were analyzed for the association of uORFs with lower average TEs.

Thus, it's important to determine if the effect is even stronger when considering only medium/high-confidence translated uORFs.

We have performed these analyses with only uORFs designated as medium/high confidence (Supplemental Figure S1I&J), and observe no clear difference between the classification categories. Selecting only transcripts with medium/high confidence translated ORFs will impose a detection bias, in that it will also tend to select for those with a higher mRNA expression. Conversely, low confidence uORFs are often low confidence due to lower mRNA expression and/or low phasing caused by overlapping uORFs, rather than lower ribosome density.

-p. 6, bottom para.: it's unclear whether the uORFs identified in the human and mouse cell datasets were subjected to the same analysis of reading frame phasing using only 28-29 nt reads to identify high/medium confidence uORFs, whether a requirement only for a peak at the AUG and stop codon was imposed, or whether all predicted uORFs were analyzed without any attempt to obtain evidence for their being translated. Much more detail is required in the description of how they did their analyses of these published datasets.

We have added additional language to the manuscript and methods section to clarify the treatment of mammalian uORFs. Because of the differences in phasing, sequencing, and RPF treatment in the human and mouse samples we did not run the ORFScore analysis on these samples to distinguish between high or low confidence uORFS in mammals.

-last sentence of p. 6, and lines 6-7 on p.7: the authors should be more careful in interpreting the association between the presence of uORFs and reduced average TEs for different gene sets. This association is consistent with uORFs exerting translational repression but certainly does not demonstrate it in the manner implied in these two sentences. The last sentence in the section on uORF abundance correlating with mRNA levels on p. 7 is similarly too strong.

We have adjusted the language here and in other sections to reduce unsupported implications of direct repression / causality. Additionally, we have improved the analysis linking translation to repression, we have added many additional cell/tissue types and show that across samples, uORF translation is linked to repression (Fig. 3E). Most importantly, we have conducted reporter experiments (now included in Fig. 5) that clearly demonstrate the repressive effect of uORFs/oORFs in various configurations and add support for causal repression.

-Fig. 3E: It seems important to use a statistical test to establish that the association of uORFs with lower TE values in embryoid bodies is significantly weaker than the correlations observed in the other mouse or human cell lines shown in Fig. 3 and Fig. S2. Also, in lines 14-16 on p. 7, the results in Fig. 3E-F should be compared to all others in Fig. 3 and Fig. S2 to make this point.

-Fig. 3E: It also seems important to verify that the TE of the set of uORFs they identified using their uORF-discovery approaches in embryoid bodies is actually lower than the TE of uORFs identified in other mouse or human cells, rather than just citing this fact from Ingolia et al (2011) as it's unclear how much overlap there might be between the uORFs interrogated in Ingolia et al (2011) and those examined here.

-The embryoid data is being used to establish that the level of uORF translation is inversely related to the TE of the CDS. A more compelling argument could be made by identifying zebrafish uORFs that exhibit changes in TE during development and determine whether the downstream CDS exhibit inverse changes in TE in the manner expected if uORFs are generally repressive. If uORFs generally don't change their TEs during development, then they should ask whether there is an inverse correlation between the TE of the uORF and the TE of the downstream CDS in different zebrafish, mouse and human cells, as it is not the number of uORFs but rather the TEs of the uORFs that should most strongly determine how much they repress downstream initiation by diverting scanning ribosomes from reaching the AUG for the CDS. They could also ask whether there is an inverse correlation between the changes in TE in uORFs and changes in TE in the downstream CDS when comparing mouse embryoids to MEFs. Similarly, the conclusion that uORFs generally decrease mRNA stability would be more convincing if changes in uORF TEs during development were inversely correlated with changes in mRNA levels, or if the TE of uORFs (not just uORF number) was inversely correlated with mRNA levels. Also, if uORF translation evokes reduced mRNA

levels, then the negative correlation between uORFs and RNA levels shown in Fig. 2G should be significantly less apparent in mouse embryoid bodies vs mESCs or MEFs.

Thank you for pointing out caveats to the EB analysis. After further consideration and addition of many more datasets, we have replaced the embryoid body section with a more thorough population level analysis over the set of all uORFs and CDSs in many samples and cell/tissue types (this can be found in Figure 3E) where we link uORF translation levels to translational repression. In regards to measuring the variability of uORF translation over developmental stages at the level of individual transcripts/ORFs, because of their short size (2 to ~20 codons), we have observed that tracking the translation efficiency of individual uORFs between samples does not have the same level of accuracy that for longer CDSs - in previous analyses we observed that splitting long CDS into short fragments (5-10AA) results in a significant level of noise even within the same CDS. For that reason, and because interactions between uORFs could cause changes in the translation of individual uORFs, we evaluated these effects over the population of transcripts across samples. The data presented now in Figure 3 analyzes ribosome footprinting data across 18 human and mouse samples and shows a correlation between the translation of the uORF and the level of repression of the CDS.

-last 2 lines of p. 7: this is confusing. Fig. S4 shows no correlation between TLS length and the CDS TE for mRNAs lacking uORFs. Presumably, this allows them to rule out the possibility that mRNAs with more uORFs have a lower TE because they tend to have longer 5'UTRs/TLSs. Is the latter true, ie. do mRNAs with more uORFs tend to have longer TLSs?.

mRNAs with more uORFs indeed tend to have longer 5'TLSs (Pearson's r=0.82), we have included this in supplementary figure S4.

-p. 8, first full para.: It's hard to be convinced that oORFs are associated with lower TEs than uORFs because they have grouped together transcripts with 1 or 2 oORFs and are comparing the results to transcripts with 1, 2, or 3 uORFs. The proper comparison would be to transcripts with 1 or 2 uORFs, with a statistical test to show a stronger anticorrelation for the oORFs vs. uORFs.

We have now included in the main text a direct comparison of transcripts with 1 uORF vs. those with 1 oORF (section: uORFs are associated with widespread…) and find a statistically significant difference between the two (5hpf, Wilcoxon p=1.23e-3). This is also demonstrated in our reporter analysis shown in Figure 5 where one oORF is more repressive than 1 uORF.

-p. 8 and Fig. S5B: The correlation shown in Fig. S5B is very small in magnitude even if highly significant, questioning the biological importance of intercistronic distance. Given the strong theoretical basis for the effect of intercistronic distance on reinitiation from the work of Kozak and Hinnebusch/Wek (on GCN4/ATF4), this analysis should be repeated with the high/medium-confidence translated uORFs; and they should also ask whether intercistronic distance is directly correlated with the TE of the uORF as would be expected if large distances are being selected for during evolution to mitigate the inhibitory effects of well translated uORFs by allowing high levels of reinitiation.

Thank you for suggesting this additional analysis. Indeed, uORF TE is inversely correlated with uORF length (r=-0.22, p=1.0e-92, 5hpf), however we do not observe significant correlation between intercistronic distance and uORF TE (r=0.002, p=0.83, 5hpf). These results have been included in the manuscript.

-p. 8 text and Fig. S5C: Given the lack of correlation between uORF length and TE of the CDS, it also seems important to re-do this analysis considering only the high-/medium-confidence uORFs, and also to ask whether uORF length is inversely correlated with the TE of the uORF as would be expected if long uORF length is being selected against in well translated uORFs to permit higher levels of reinitiation.

Performing this analysis with only the high/medium confidence ORFs (which does impose bias on the uORF lengths) still reveals no significant correlation. We have performed the suggested analysis to check the correlation of uORF length and uORF TE and observe that indeed, uORF length is

inversely correlated with uORF translation efficiency (Pearson's r=-0.31, p=2.49e-31) – this result has now been included in figure S5E.

-Fig. S5: Although there is a significant negative correlation here, it is quite weak. It should be verified that this is true even if high-/medium-confidence translated uORFs are analyzed.

In this case, figure S5D (the panel with significant negative correlation) shows AUG score versus translation efficiency. By selecting only the subset of high/medium-confidence translated uORFs in single-uORF transcripts for this analysis, we would impose a significant bias since high/medium confidence ORFs are by nature translated and AUG context drives that translation. Rather, we elected to show the correlation over the whole range of both translation and AUG score for this figure.

-p. 9, top para. and Fig. 4A: Given that the ORF AUG score is so much lower for oORFs versus uORFs, yet oORFs appear to be more inhibitory, they should determine whether the TE of oORFs is also generally lower than that of uORFs. If so, the relatively greater inhibition by oORFs, despite being translated by fewer scanning ribosomes, would derive from the fact that they overlap the main CDS in a way that precludes reinitiation and also might interfere with initiation complex assembly at the CDS AUG codon.

We agree with the reviewer that it is likely that oORF are stronger repressors because they likely block re-initiation. This is supported by our reporter assay, since the only difference between both constructs is the absence of a stop codon in the oORF. On the other hand, it is technically difficult to measure oORF translation, because the CDS start peak and out-of-frame reads will greatly bias the reported TE for oORFs, such that their measured 'translation' will be affected by the CDS translation, biasing further analyses. Though it could be possible to use just the non-overlapping region to measure translation, these regions are very short and too susceptible to noise to quantitatively compare their translation to uORFs. Nevertheless, we address this indirectly in that the initiation contexts of oORFs are significantly worse than uORFs, which should result in less oORF translation and thus buffer their repressive effects as suggested.

-Fig. 4C: the schematics are hard to fathom and should be explained in the legend.

We have removed figure 4C as it did not add any significant information, and have instead clarified the text.

-Fig. 4F and text on p. 9: Again it's unclear whether all potential uORFs are examined for mean length, and whether the conclusion also holds for high/medium confidence translated uORFs, and also whether the TE of the uORF is inversely correlated with its length as might be expected if short uORF lengths are being selected during evolution for well-translated uORFs?

Thank you for suggesting this additional analysis. Indeed, uORF TE is inversely correlated with uORF length (r=-0.22, p=1.0e-92, 5hpf) and this has now been included in the text.

-It's not explained where the cited figure of <60nt (presumably 41.5 nt) for the mean uORF length in all vertebrate species comes from-which species, which types of uORFs considered? They also don't indicate the mean uORF length obtained in the TLS shuffle simulation-how much larger than 41.5 nt is it?. Finally they don't discuss any mechanistic basis for the apparent selection pressure against uORF length.

We do include all uORFs in this analysis since even those that are not translated are potentially subject to selective pressures on their sequence and presence if their expression and translation in some condition or cell type is detrimental to the organism. We have also included language to clarify that we are visualizing all ORFs, added an additional supplementary figure (Figure S9) displaying the distributions and means in human and mouse, and indicated the mean length from the shuffling simulations in each plot. Selective pressures acting upon uORF length likely act on the emergence and loss of start and stop codons within the 5'TLS, an effect which could potentially be studied in further depth by examining rates of codon gain/loss between organisms but that we feel is outside of the scope for this paper.

-bottom of p. 9/top of p. 10: cite Fig. 4D versus 4C.

Thank you, we have corrected this citation.

-p. 10, line 5: context data for human/mouse uORFs should be shown in Supplemental Figs.

Supplementary panels display the distributions of ORF context scores in humans/mice in supplementary figure S6B,C.

-Fig. 5B-D: first, this figure was not cited in the text.

This figure has now been cited in the text.

Second, I assume that log(2) ratios are being plotted.

We have added to the figure legend and in the axis to indicate that log2 ratios are being plotted.

Third, it is surprising that the ~average UTR/CDS TE ratio generally appears to be ~2, indicating that there are typically 4-fold greater ribosome densities in 5'UTRs vs.CDS. This would be surprising, considering the typical genes shown in Fig. 1D where uORF densities are less then CDS densities, and the fact that most genes don't contain multiple uORFs and uORF AUG context is so poor.

Though uORF AUG contexts are unfavorable compared to the CDS, they are still the first AUG encountered by the ribosome, and thus will still be engaged with some baseline frequency. Additionally, many 5'TLSs are short, and the RPF contribution from multiple uORF initiation and termination peaks contributes more to the translation signal when calculating the TE of a short 5'TLS as opposed to the translation within the internal codons of a CDS, this effect should not affect comparisons between samples.

-Fig. 5B-D and text on p. 10, 1st full para. It's not clear that the differences in correlation coefficients between panels B, C,D are significantly different from one another. Perhaps a better analysis would be to make notched box-plots of the ratios in the different cell types and determine whether the notches fail to overlap for the comparisons in C and D but not for panel B.

We have changed this figure after the addition of additional samples, such that we can now directly compare the same tissue/cell types between human and mouse. We feel that the difference between the background condition (shuffled pairings) and orthology pairings is now clearer.

-p. 10, line 6 from bottom: stipulate what kind of uORFs is referred to here-I believe it's high/medium confidence translated uORFs, but not sure for the oORFs. Also what kind of uORFs is referred to a few lines down in citing 61% of zebrafish uORFs being translated?

We have added language to the manuscript clarifying which type of ORFs are referred to for these statistics (The first statement makes no implications of translation, whereas the second only includes low/med/high confidence translated uORFs).

-p. 12, lnes 12-14: This is an overstatement. They have shown a reasonably strong correlation between the relative TEs of uORFs versus downstream CDSs across species. This is consistent with, but certainly does not prove that regulation by particular uORFs is conserved across species. A few lines down, it is also incorrect that uORFs have sequence contexts less favorable than by chance, which is true only of oORFs.

We have adjusted the language to reflect the nature of the results. As for the sequence contexts, after reviewer 3's suggestion to replace CAU context with a more unbiased background estimation, we used a random sample of 50,000 nucleotide contexts from the 5TLSs of each respective organism - we observe that the sequence contexts surrounding uORF initiation codons are indeed less favorable than by chance (Figs. 4A, S6A-C).

Referee #3:

Specific major concerns
1. Metagene analysis can be heavily biased by genes with a particularly strong signature- a distribution of values compiled for the metagene scores is a useful supplement to confirm the effect isn't driven by a single, large effect gene.

We normalize metagene profiles per-gene as well as across the population, so that each gene contributes an equal amount to the metagene plot, which should mitigate contribution by strongly translated ORFs. We have added language to the methods section to clarify this.

2. The authors describe TE/repression as difficult to study in cases of oORFs because of overlapping regions/counts. This is can be circumvented using computationally defined boundaries that exclude the region of overlap: for the CDS downstream the of the stop codon in the oORF, and for the oORF, upstream of the CDS. Moreover, in cases where the CDS and oORF do not share a frame, information can be gained even from this region of overlap by looking at the framing data itself rather than regional counts.

It is technically difficult to measure oORF translation, because the CDS start peak and out-of-frame reads will greatly bias the reported TE for oORFs, such that their measured 'translation' will be affected by the CDS translation, biasing further analyses. Though it could be possible to use just the non-overlapping region to measure translation as suggested, these regions are very short and too susceptible to noise to quantitatively compare their translation to uORFs. Nevertheless, we address this indirectly in that the initiation contexts of oORFs are significantly worse than uORFs, which should result in less oORF translation and thus buffer their repressive effects.

3. The use of CAU context versus AUG context for TLS background estimation

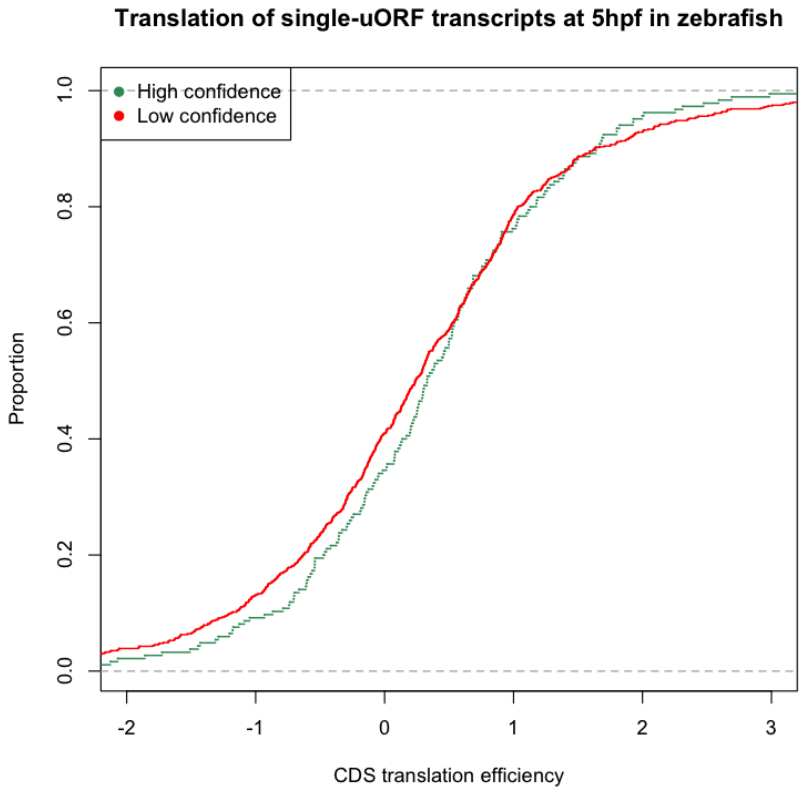
Thank you for pointing this out. We have replaced CAU context with a more unbiased radom sample of 50,000 random nucleotide contexts from the 5TLSs of each respective organism.

Minor concerns
5. Per the reliance of this paper on several previously published papers (e.g. ORFScore, PhyloCSF), a brief description of them would be of considerable value for assaying the merit of the work.

We have added more detailed description of the analytical methods/tools to the manuscript.

6. In 4B, there is potential for ascertainment bias - a good AUG context may confer better translation, and this may in turn better regulate downstream CDS TE, but also ensures higher confidence detection of the uORF. It would be useful to compare AUG context in high versus low confidence uORFs, and then compare how these affect downstream TE.

We present the differences in context between the different confidence level ORFs in figure S6A, and indeed observe the expected direction of difference in AUG context (lower confidence uORFs tend to have less favorable contexts), though there is no statistical difference in the translation of the CDS in zebrafish transcripts with a single high-confidence uORF vs. a single low confidence uORF (Wilcoxon p=0.26, plot attached below). It is worth noting that low confidence uORFs might still be translated but not confidently detected because of low RNA levels or overlapping uORFs which influences the phase observed in either of the uORFs being analyzed.



7. 4D isn't discussed in the text. There is a discussion of potentially similar data, but refers to an unrelated cartoon figure 4C.

We have corrected this citation.

8. Several issues in figure or legend clarity:
a. 1A lacks adequate labeling per color coding (orange v grey v purple) and has some difficult to distinguish colors (low v med v high) and ribosomes (med v high). The legend is also unclear/misleading in describing a subset of uORFs as containing oORFs.

Colors have been corrected on the figure to give more visual distinction between confidence levels, and we have removed the unclear language about overlapping ORFs.

b. the arrangement of 1C is non intuitive given the way these are usually presented (left to right as the standard vs the up to down for 5' to 3'). Moreover, the framing data in Low Confidence uORFs is described as potentially due to overlapping frames, though there's no clear evidence in this figure to support that.

We have adjusted the arrangement of 1C so that metaplots for each ORF type are side-by-side to match the standard presentation. The language referring to ORFs overlapping low confidence ORFs has been changed in the text. Mainly the evidence to support overlapping uORFs is that after the stop codon we observe phasing in the frame3 which is not observed for the other examples

c. 1D is nigh illegible because of size; moreover, the CDS isn't depicted in a graphically meaningful way. This makes it difficult to judge the size, location, and relevance of the uORFs. Additionally, "mRNA Seq" or RPKM is a more informative title axis than simply "Input".

We have added a black border to CDS ORFs and changed the lower Y-axis label to "RNA-seq" for visual clarity.

d. fig 3 changes in axes change potential magnitude of effect

e. 3C,D have no definition of LIF in the figure, supplement, or text-if the absence or presence is worth noting, it is an acronym worth defining.

We have added text clarifying this, as well as a supplementary table that directly maps each sample in the paper to the source data and repositories used from public databases with information about each sample.

This figure also uses a different x axis than 3A,B,E,F. Changes in the x-axis on cumulative distributions can be misleading for magnitude of effect. If all six graphs are to be compared to each other to remark on differences or similarities between species, the axes need to be the same.

Thank you for pointing this out, we have adjusted the x-axis scales of all panels in Figure 3 to be the same, so that the plots can all be visually compared.

f. 4A's x axes is not intuitively clear (how you would grade a better context) without having read Grzegorski.

We have added language to the legend, text, and methods section to clarify that these are nucleotide frequency matrices that have been validated in vivo in zebrafish.

g. 4C is a graphical abstract missing enough pieces of data to actually clarify what it is intending to show.

We have removed figure 4C as it did not add any significant information to the story, and have instead clarified the text.

h. Figure 5E has r=.237 and is considered no correlation, whereas S5B has r=.147 and is considered correlated.

We have changed this figure after the addition of additional samples, such that we can now directly compare tissue/cell types between human and mouse.
Addressing the original figure, though 5C&E (formerly 5E) are used in comparison to more correlated sample pairs, we do not consider them uncorrelated, rather this control was performed to establish the baseline/background level of correlation that arises from the concordance of CDS translation across species (if 5'UTRs were not translated at all, the denominator of the 5UTR-CDS ratio (CDS translation) will still be correlated to some degree). Thus, within Figure 5, though there is some baseline level of correlation due to CDS translation, we observe a much higher level when taking into account the specific relationship between 5'TLS and CDS translation.

i. S5C,D lacks legend descriptions

We have now included proper legends for this and all other supplemental figures.

2nd Editorial Decision 04 January 2016

Thank you again for submitting the revised version of your manuscript and sorry for the slightly extended duration of the re-review period. Your study has now been seen by two of the original referees whose comments are shown below. As you will see they both find that all major criticisms have been sufficiently addressed, although ref #2 still points to a few sections where further analysis/clarification is needed. I would therefore ask you to submit a final revision of the work in which you address the remaining minor concerns from the referees as well as the following editorial points:

-> Please make sure all figures are referenced in the manuscript text. Figs S8 and S8 are currently not referenced (as also pointed out by ref #2) and fig S10 is only referenced in the legend to figure 5. In addition, I have to ask you to restructure and relabel the supplemental figures (both in figure panels, legends and manuscript text) according to our author guidelines for expanded view (http://emboj.embopress.org/authorguide). Given the large number of supplemental files in this case, I would recommend that you provide everything as a single appendix file pdf and use the

corresponding nomenclature. The supplemental tables should be relabeled as Expanded View tables. Please feel free to contact me with any questions on this.

-> Please include GEO links for sequencing data as soon as this becomes available.

-> Please also make sure that all figure legends contain information on statistical test employed, the nature of the error bars and the underlying number of replicas whenever relevant.

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal, I look forward to your final revision.

REFEREE REPORTS

Referee #2:

The authors have made significant improvements to the paper, but a few issues still remain:

-The new figure 3E purports to show that the translational repression of the CDS is correlated with the TE of the uORF. While the correlation is significant when all of the different datasets are included, it seems likely that it would not be the case if the zebrafish data was analyzed separately. This is rather troubling as these data are a focal point of the study. The zebrafish data should be analyzed separately for all uORFs, and also for only the high- and medium-confidence uORFs, and the results should be presented and discussed.

-p. 7, top: Fig. S1I is mistakenly cited as Fig. S2J; this was done on p. 8, para. 1 as well for the citation to Fig. S1J; and lower on p. 8 for Fig. S1B,D,F,H

-Fig. S2: no P-values were provided to document the significance of these effects. The legend should at least state that all P-values were below a certain threshold value.

-p. 8, 2nd and 3rd paragraphs: figure citations in both of these paragraphs are incorrect or inadequate.

-Fig. S5B: while significant, this correlation is surprisingly weak. This analysis could be repeated for the high- and medium-confidence uORFs and the results described in the text, as the results in Fig. S5B imply that intercistronic distance makes a surprisingly small contribution to repression by uORFs.

-p. 9: The sentence "we did not observe any significant effect of uORF length on CDS translation efficiency (Fig S5C)." should be amended to include the clause "even when the analysis was restricted to high- and medium-confidence uORFs.

-Neither Fig. S6A nor S6D were cited in text, but should be on p. 9 in relation to Fig. 4A. The sentence on p. 9 that I think refers to Fig. S6D cites an r value of 0.22 versus that shown in the figure of 0.3. Which is correct?

-The analysis of relationship between uORF TE and uORF AUG context score shown in Fig. S6D could be repeated with high- and medium-confidence uORFs, as it seems surprisingly weak. Based on the findings in Fig. S6A, one might expect that the correlation will be much stronger for the subset of high- and medium-confidence uORFs.

-p. 10: the meaning of this sentence is unclear, much less the methods employed to conduct the analysis: "The depletion/enrichment of uORFs per gene was significantly correlated between mammalian one-to-one homologs (Pearson's $r=0.39$, $p<2.2e-16$)."

Referee #3:

The revised manuscript has addressed our major concerns with the original submission and the new data added, especially the uORF reporters, strengthen the paper.

As a minor concern, language suggesting causal relationships remains in parts of the paper. For example, p. 8, "we investigated the specific features that determine the level of repression conferred by uORFs" but the analysis here is correlative. It remains important to be cautious in these genome-scale data, because selection for low protein abundance could independently act through reduced mRNA level, reduced ribosome recruitment (independent of uORF-mediated repression), and repressive uORFs.

2nd Revision - authors' response 07 January 2016

Thank you for the quick turnaround over the holiday season, and again to the editor(s) and all reviewers for all of the helpful feedback, it has been very useful in improving our manuscript. We have reformatted the manuscript as requested and addressed the points made below and hope that the manuscript is now suitable for publication.

-> Please make sure all figures are referenced in the manuscript text. Figs S8 and S8 are currently not referenced (as also pointed out by ref #2) and fig S10 is only referenced in the legend to figure 5. In addition, I have to ask you to restructure and relabel the supplemental figures (both in figure panels, legends and manuscript text) according to our author guidelines for expanded view (http://emboj.embopress.org/authorguide). Given the large number of supplemental files in this case, I would recommend that you provide everything as a single appendix file pdf and use the corresponding nomenclature. The supplemental tables should be relabeled as Expanded View tables. Please feel free to contact me with any questions on this.

We have merged and reorganized supplemental figures such that the manuscript now contains 5 Expanded View figures and 3 appendix-supplemental figures. All figures, legends, and tables have been relabeled and included according to the author guidelines provided above.

-> Please include GEO links for sequencing data as soon as this becomes available.

We have now included the SRA repository number for the deposited sequence data

-> Please also make sure that all figure legends contain information on statistical test employed, the nature of the error bars and the underlying number of replicas whenever relevant.

Figure legends and p-values in the text now contain information on which statistical test was employed, and error bars/replicates are stated clearly wherever relevant.

-> Papers published in The EMBO Journal include a 'Synopsis' to further enhance discoverability. Synopses are displayed on the html version of the paper and are freely accessible to all readers. The synopsis includes a short standfirst - written by the handling editor - as well as 2-5 one sentence bullet points that summarise the paper and are provided by the authors. I would therefore ask you to include your suggestions for bullet points

We have included bullet point and standfirst suggestions in the file 'standfirst and bullet points.docx'

Referee #2:

The authors have made significant improvements to the paper, but a few issues still remain:

-The new figure 3E purports to show that the translational repression of the CDS is correlated with the TE of the uORF. While the correlation is significant when all of the different datasets are included, it seems likely that it would not be the case if the zebrafish data was analyzed separately.

This is rather troubling as these data are a focal point of the study. The zebrafish data should be analyzed separately for all uORFs, and also for only the high- and medium-confidence uORFs, and the results should be presented and discussed.

The power of this particular analysis lies in being able to compare and combine results across samples and conditions, and it would be difficult to make a meaningful conclusion from only five zebrafish samples alone. Furthermore, the zebrafish ribosome profiling is whole-embryo, thus translation effects may be compensated for between cell types and less clear in the aggregate signal for this particular analysis. Reducing this analysis to only high/medium confidence uORFs would further reduce this power and introduce biases (such as RNA expression; we note that the confidence levels refer to confidence of detection, not level of translation) and do not think this would substantially alter the conclusions of the paper.

-p. 7, top: Fig. S1I is mistakenly cited as Fig. S2J; this was done on p. 8, para. 1 as well for the citation to Fig. S1J; and lower on p. 8 for Fig. S1B,D,F,H

These figure calls have been corrected in the manuscript

-Fig. S2: no P-values were provided to document the significance of these effects. The legend should at least state that all P-values were below a certain threshold value.

P-values have been calculated, and a note added to the legend that p-values for all comparisons shown are highly significant ($p<0.0005$, although most sample comparisons are orders of magnitude lower than this threshold).

-p. 8, 2nd and 3rd paragraphs: figure citations in both of these paragraphs are incorrect or inadequate.

We have corrected the mistaken figure citations in this section

-Fig. S5B: while significant, this correlation is surprisingly weak. This analysis could be repeated for the high- and medium-confidence uORFs and the results described in the text, as the results in Fig. S5B imply that intercistronic distance makes a surprisingly small contribution to repression by uORFs.

We have repeated this analysis with just high or medium confidence translated uORFs and observe that the inverse correlation between intercistronic distance and CDS TE is slightly higher when selecting only translated uORFs. This has been included in the text.

-p. 9: The sentence "we did not observe any significant effect of uORF length on CDS translation efficiency (Fig S5C)." should be amended to include the clause "even when the analysis was restricted to high- and medium-confidence uORFs.

We have added this clarification to the text

-Neither Fig. S6A nor S6D were cited in text, but should be on p. 9 in relation to Fig. 4A. The sentence on p. 9 that I think refers to Fig. S6D cites an r value of 0.22 versus that shown in the figure of 0.3. Which is correct?

We have added the proper supplemental figure citations to this section of the text. The r-value of 0.22 refers to high confidence uORFs, whereas the r-value of 0.3 refers to all uORFs in single-uORF transcripts, this has been clarified in the text.

-The analysis of relationship between uORF TE and uORF AUG context score shown in Fig. S6D could be repeated with high- and medium-confidence uORFs, as it seems surprisingly weak. Based on the findings in Fig. S6A, one might expect that the correlation will be much stronger for the subset of high- and medium-confidence uORFs.

We have repeated this analysis for the high- and medium- confidence uORFs (in single uORF transcripts to avoid the confounding effect of multiple uORFs acting on each other), comparing

these to all single uORF transcripts, and while selecting the subset of high and medium confidence uORFs biases them towards higher translation efficiency, the correlation of AUG score and uORF translation is similar to all uORFs: all single uORF transcript uORFs r=0.26, high confidence uORFs r=0.22, medium confidence uORFs r=0.28. Given that uORF TE measurements tend to have higher noise than CDS regions due to their size, and the AUG scoring scheme we use is based on mononucleotide frequencies, r~0.30 is within our expectations as there are other factors besides initiation sequence context which contribute to translation efficiency.

-p. 10: the meaning of this sentence is unclear, much less the methods employed to conduct the analysis: "The depletion/enrichment of uORFs per gene was significantly correlated between mammalian one-to-one homologs (Pearson's r=0.39, p<2.2e-16)."

Thank you for pointing this out, we have added language clarifying that this refers to a correlation of the uORF enrichment z-scores across TLSs in the homology set used for the later conservation analysis.

Referee #3:

The revised manuscript has addressed our major concerns with the original submission and the new data added, especially the uORF reporters, strengthen the paper.

As a minor concern, language suggesting causal relationships remains in parts of the paper. For example, p. 8, "we investigated the specific features that determine the level of repression conferred by uORFs" but the analysis here is correlative. It remains important to be cautious in these genome-scale data, because selection for low protein abundance could independently act through reduced mRNA level, reduced ribosome recruitment (independent of uORF-mediated repression), and repressive uORFs.

Thank you, we have corrected this and gone through the manuscript to remove any other unsupported statements of causality.

